

Monocular Visual Odometry with Unmanned Underwater Vehicle Using Low Cost Sensors

Original

Monocular Visual Odometry with Unmanned Underwater Vehicle Using Low Cost Sensors / Dabove, P.; DI Pietra, V.; Piras, M.. - STAMPA. - (2020), pp. 810-816. (Intervento presentato al convegno 2020 IEEE/ION Position, Location and Navigation Symposium, PLANS 2020 tenutosi a Portland (OR - USA) nel 2020) [10.1109/PLANS46316.2020.9109841].

Availability:

This version is available at: 11583/2838377 since: 2020-07-06T08:53:27Z

Publisher:

Institute of Electrical and Electronics Engineers Inc.

Published

DOI:10.1109/PLANS46316.2020.9109841

Terms of use:

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

IEEE postprint/Author's Accepted Manuscript

©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

Monocular visual odometry with unmanned underwater vehicle using low cost sensors

Paolo Dabove, Vincenzo Di Pietra, Marco Piras

Department of Environment, Land, and Infrastructure Engineering
Politecnico di Torino

Corso Duca degli Abruzzi 24, 10129, Turin (Italy)
(paolo.dabove, vincenzo.dipietra, marco.piras) @ polito.it

Abstract— The positioning in underwater environments is today a strong necessity for many purposes, such as construction, communication, localization, and environmental monitoring. The use of underwater rovers allows to perform visual inspections, maintenance and repair of many infrastructures, like dams, pipes, tunnels, structures as well as the analyses of the underwater environments in lakes, rivers and seas. This work deals with interesting results about monocular visual odometry for unmanned underwater navigation systems. Interesting results have been obtained considering low-cost sensors simulating the real-time, challenging operational conditions, applied for a dedicated archaeological situation. Particular algorithms for navigation procedures and outliers' rejections have been written and will be presented in this paper: these aspects have a great importance especially for autonomous navigation solutions in underwater complex environments, such as for archaeological applications.

Keywords—underwater positioning; visual odometry; low cost; autonomous navigation; positioning.

I. INTRODUCTION

Nowadays, there are several strong motivations to replace some dangerous human activities with the use of small unmanned terrestrial, aerial and underwater systems. An interesting example is the use of unmanned underwater vehicles (UUVs) in particular environments, such as archaeological ones, or for other dedicated activities like inspections.

UUVs are fundamental tools for several emerging applications which tasks must be accomplished underwater. The capability of these systems to reach critical areas and provide real-time video frames of the neighborhood is fundamental in several research and industrial applications like photogrammetry, inspection, service robotics and more. Compared with other systems remotely piloted, like unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs), UUVs presents several challenges both in terms of image acquisition and in terms of localization and navigation. First of all, the global navigation satellite system (GNSS) positioning is not available underwater, as the signal can't reach far below the water level. Then, the dynamic motion of the system is quite complex with respect to pedestrian or ground vehicle movements, mainly due to the unstable

underwater environment. Alternative positioning methods such as acoustic positioning systems and inertial navigation are preferred when the vehicle navigates in easy environments. In these cases, the main problems are due to the noise of observations and the drift of the final solution. Thus, a better possible alternative could be to integrate several sensors on the unmanned system, where the positioning solution is obtained with a fusion of different sources of data, like images, ranges, inertial observations as accelerations and angular velocities, pressure, or techniques like object detection, etc.

The problem of estimating a vehicle's motion from visual input alone started in the early 1980s and was described by Moravec [1]. Over the years, monocular and stereo VOs have almost progressed in order to reach a positioning level of accuracy of about few tens of decimeter. This is deeply investigated for terrestrial navigation, while this is not happened for underwater environments, where it is still challenging. In this context, the Authors have investigated and developed an innovative method for underwater navigation considering low-cost sensors. The algorithms, implemented in Matlab, start to the acquisition up to the processing phases. Particular attention was paid to features extraction: this is the first step in any image analysis procedure and it is essential for many applications. As described in literature, there are two main approaches to find feature points and their correspondences: the first one is to find features in one image and track them in the following images using local search techniques, such as correlation. The second one, used in this research activity, is to independently detect features in all the images and match them based on some similarity metric between their descriptors. During the feature-detection step, the image is searched for relevant keypoints that are likely to match well in other images. A local feature is an image pattern that differs from its immediate neighborhood in terms of intensity, color, and texture. For Visual Odometry (VO), point detectors, such as corners or blobs, are important because their position in the image can be measured accurately.

The positioning solution has been obtained considering images and inertial measurements as input of a Kalman filter approach coupled with a data snooping technique, without using any external sensors (e.g., pressure information). This research was made considering an archaeological application, where the positioning accuracy and level of details are

particularly important. The results have shown an impressive performance in terms of positioning precision and we believe that these results will have a great impact especially for future underwater navigation solutions.

II. UNDERWATER POSITIONING

Considering the underwater positioning technique, one of the main problems is to localize the user without using the Global Positioning System (GPS). This problem can be overcome coupling high-grade inertial measurement units (IMUs) with other sensors, like compass and pressure sensor [2]. In this case, the common approach is defined as dead-reckoning localization but, as it is widely described in literature, the final results are strongly dependent by the quality of the sensors employed. In order to overcome these issues, simultaneous localization and mapping (SLAM) techniques are widely used: these methods consider many other sensors, like Lidar, sonar and cameras, whose don't suffer from drifts and their results are time-independent in terms of biases. Starting from the last decade, two main techniques have been developed: the so-called visual SLAM (VSLAM) and the visual-inertial odometry (VIO) [3]. If in the first case the solution is obtained considering only images, in the second one the final solution is estimated adding also other sensors, like the inertial ones. Since sometimes it is not possible to have more than one camera installed on underwater Remotely Operated Vehicles (ROVs), some researchers have been developed algorithms that consider only one camera, bringing the concept of the monocular visual odometry from terrestrial applications [4] to the underwater environment [5][6][7]. The possibility to add information from inertial sensors [8], has allowed to obtain impressive results, reaching an accuracy of less than 10 cm in terms of localization [9][10], also thanks to the diffusion of public datasets both for terrestrial [11][12], aerial [13] and underwater environments [14][15][16].

In this last case, many factors affect the quality of the final results, as the quality of sensors used and the environmental conditions, like water turbidity, particles suspensions, backscattering, reflections and poor illuminations. For these reasons, it is really challenging to obtain an accurate solution, especially if low-cost sensors cameras are used. In literature some studies related to the investigation of monocular cameras are available [17][18] especially if coupled to low-cost IMUs and pressure sensors [19] or sonar [20]. In addition, most of public datasets available today don't provide basic information for processing the images, such as the camera parameters (e.g., the focal length, distortions) and all sensors' calibration. For this reason, in this paper part of the AQUALOC dataset [16] (freely available from <http://www.lirmm.fr/aqualoc/>) has been used, in order to develop an algorithm based only on monocular visual odometry which allows to obtain an accurate positioning method in underwater environment.

III. MATERIALS AND METHODS

The AQUALOC dataset has been collected by French colleagues with the aims of developing new VSLAM and VIO positioning techniques especially for underwater environment. This dataset composed by several measurement campaigns and acquisition systems, mainly based on a monochromatic camera,

a microelectromechanical system (MEMS)-based IMU, a pressure sensor, and a computing unit for synchronous recordings. The dataset comprises data collected in three different test-sites, a harbor and two different places more related to archaeological sites. In this paper only one dataset is considered and its characteristics are summarized in Table 1. This choice is due to the environmental conditions, which are really challenging for a monocular visual odometry approach.

TABLE I. DETAILS OF THE DATASET CONSIDERED (FROM AQUALOC SEQUENCES) AND ITS VISUAL DISTURBANCES

| Site | Sequence | Duration | Length | Depth |
|-------------------------------------|----------|----------|--------|---------|
| 2 nd archaeological site | #4 | 11'09" | 18.1 m | ~ 380 m |

In the considered case, many visual disturbances are present, like turbidity, backscattering, sandy clouds and turbulence due to the robotic arms.

Even if the rover is composed by a monochromatic camera, a pressure sensor, a low-cost MEMS-IMU, and an embedded computer, in this paper only images will be considered. As visible from Fig. 1, the camera is placed behind an acrylic dome to minimize the distortion effects induced by the difference between water and air refractive indices. The image acquisition rate is 20 Hz. The IMU delivers measurements from a three-axis accelerometer, three-axis gyroscope, and three-axis magnetometer at 200 Hz. The embedded computer is a Jetson TX2 running Ubuntu 16.04 and is used to synchronize all measurements collected by all different sensors. The micro-computer is equipped also with an SSD to store the sensors measurements.

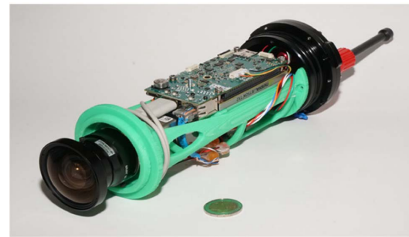


Fig. 1. The acquisition system considered in this work (courtesy of [16])

Even if the sensors were already calibrated, we have performed another calibration using a dedicated toolbox developed at Politecnico di Torino (Italy), in order to estimate the intrinsic and extrinsic parameters for the localization. As easily found in literature [21][22][23], the camera calibration process allows to estimate the focal length (Table 2), the principal point and the distortion coefficients of the sensor used.

TABLE II. CHARACTERISTICS AND DETAILS OF THE CAMERA USED IN THESE TESTS

| Sensor | Characteristics | Detail |
|--------|-----------------|------------------|
| Camera | Model | UEye - UI-3260CP |
| | Resolution (px) | 968 x 608 |

| Sensor | Characteristics | Detail |
|--------|-------------------|------------------------|
| | Sensor | Monochromatic |
| | Frames per second | 20 fps |
| | Lens | Kowa LM6NCH C-Mount |
| | Focal length (mm) | 6 |

TABLE III. CHARACTERISTICS AND DETAILS OF THE IMU AND EMBEDDED COMPUTER USED IN THESE TESTS

| Sensor | Characteristics | Detail |
|---------------------------|--------------------------|---------------------------|
| Inertial measurement unit | Model | MEMS - MPU-9250 |
| | Gyros frequency | 200 Hz |
| | Accelerometers frequency | 200 Hz |
| | Magnetometers frequency | 200 Hz |
| Embedded Computer | Model | Nvidia - Tegra Jetson TX2 |

Even for the IMU we have performed the calibration test in order to estimate the noise model parameters: this is possible mounting the device in a leveled non-magnetic plate and logging data over 24h and subsequently to apply the Allan variance on these data [25]. The Allan variance (AV) is a well-known technique that is commonly used to identify and to quantify inertial sensors' stochastic noises, as quantization noise, random walks errors, and bias instability, among others. In applying the AV technique, it is mandatory to only process data from static measurements. Theoretical foundations about the AV can be found in the literature [26][27][28] and are beyond the scope of this paper. Since these parameters (IMU noises, camera, IMU relative transformation, and measurements time delay) are independent of the medium (air or water), it is possible to estimate them in air due to the easy operational conditions.

The test site considered in this work was located at a depth of approximately 380 meters, with environmental conditions that varies from low-textured sandy areas (Fig. 2 - left) up to places with archaeological finds, like amphorae (Fig. 2 right).

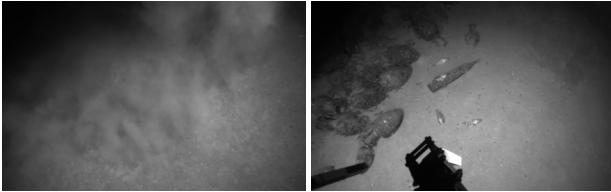


Fig. 2. The example of environmental conditions considered: a low-textured sandy area (left) and an archaeological site (right)

Considering Fig. 2, it is possible to see how some preliminary checks must be done before processing the dataset. Indeed, sometimes it is possible to find some arms of the UUV in the picture: these details create problems in the processing phase, especially for feature extraction. As it is possible to see

in Fig. 3, in presence of UUV's arms most of features extracted belong to the arm itself: this is useless for the photogrammetric approach and dangerous for the positioning point of view because it affects the final results. For this reason, it is particularly important to cut out the details related to the rover and to consider only the portion of the image related to the underwater environment (Fig. 3 – right).

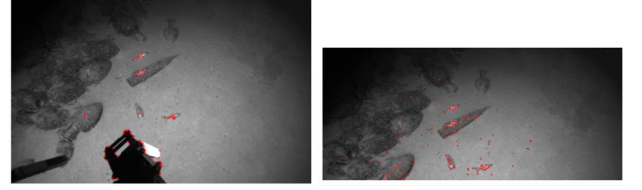


Fig. 3. Complete (left) and reshaped (right) image in presence of some UUV details

It is important to know the location where the rover is moving because the first stage of any image analysis procedure is the so-called feature extraction, where an algorithm searches in the image some relevant patterns (point detectors) whose differ from their immediate neighborhood in terms of intensity, color, and texture (feature detection phase). For visual odometry it is possible to consider many different point detectors, such as corners, edges or blobs, because their position can be measured accurately in the image.

If in presence of archaeological finds it is quite easy to extract some features due to the high variability of the image (Fig. 2 - right), in sandy areas or where the pattern has low details it is really challenging to obtain a good selection of features that allows the matching and then the positioning estimation (Fig. 2 - left). It is important to highlight that, in presence of archaeological finds, the marine wildlife is present, so even in this case there are some problems in terms of feature extractions because some fishes enter in the field of view of the camera for few seconds and then go out, increasing the noise and decreasing the number of features useful for the automatic extractions of the tie points. The simplest way for matching features between two images is to compare all feature descriptors in the first image to all other feature descriptors in the second image. Descriptors are compared using a similarity measure. Different feature detector can be considered, such as corner detectors (e.g., Forstner [29], Harris [30], Shi-Tomasi [31], and FAST [32]) and blob detectors (SURF [33] and SIFT [34]). Each detector has its own pros and cons. Corner detectors (Fig. 4) are fast to compute but are less distinctive, whereas blob detectors are more distinctive but slower to detect. Additionally, corners are better localized in image position than blobs but are less localized in scale. This means that corners cannot be redetected as often as blobs after large changes in scale and view-point.

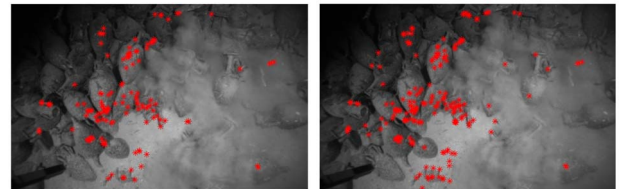


Fig. 4. Harris (left) and Shi-Tomasi (right) corner detectors

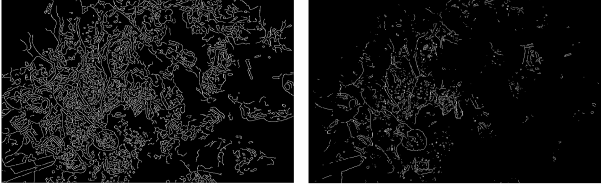


Fig. 5. Canny (left) and Sobel (right) edge detectors

However, blobs (Fig. 5) are not always the right choice in some environments. For these reasons, the choice of the appropriate feature detector should be carefully considered, depending on the computational constraints, real-time requirements, environment type, and motion baseline (i.e., how nearby images are taken). An approximate comparison of properties and performance of different corner and blob detectors is given in Table 4.

TABLE IV. COMPARISON OF FEATURE DETECTORS: PROPERTIES AND PERFORMANCE

| | Corner detector | Blob detector | Edge detector | Rotation invariant | Scale invariant | Affine invariant |
|------------|-----------------|---------------|---------------|--------------------|-----------------|------------------|
| Harris | x | | | x | | |
| Shi-Tomasi | x | | | x | | |
| FAST | x | | | x | x | |
| SIFT | | x | | x | x | x |
| SURF | | x | | x | x | x |
| Canny | | | x | x | | |
| Sobel | | | x | x | | |

For SIFT feature matching, a distance-ratio test was proposed by the authors initially, for use in place and object detection [34]. Unfortunately, the threshold for the test can only be set heuristically and it might be possible to remove correct matches as well. Thus, in some cases, it might be beneficial to skip the ratio test and let RANSAC [35] take care of the outliers, as proposed in this work. Following the approach presented in [4], in Fig. 6 it is summarized the proposed positioning technique based on SIFT detector, RANSAC outliers' rejection method and Kalman filter approach. The VO solution is obtained following different steps: after the tie points detection and extraction, the positioning solution has been obtained considering a 6-state Kalman filter approach coupled with a data snooping technique, without using any external sensor (e.g. INS, pressure). Firstly, the position of the center of view of the camera has been predicted and secondly the prediction has been corrected considering the observations into the equations. In order to strength the solution, it has decided to propagate coordinates between two consecutive epochs. The prediction of the center of view position at the generic epoch t is determined from Eq. 1:

$$x = x_0 + v_0 \cdot t + \frac{1}{2} \cdot a \cdot t^2 \quad (1)$$

considering the velocity and acceleration of the UUV:

$$v = \frac{x_{t-1} - x_{t-2}}{\Delta t} \quad (2)$$

$$a = \frac{v_{t-1} - v_{t-2}}{\Delta t} \quad (3)$$

These equations have been considered in the solution's estimation, based on the Kalman Filter. As described in [4], the solution is obtained in two different ways: for the first epoch, it has been followed the least-square (LMS) approach, considering a system of n linear equations that directly depend on n observations L_1, L_2, \dots, L_n , on the assumption that each equation contains as a term known only one quantity directly measurable, the r unknowns are X_1, X_2, \dots, X_r with $n > r$. For the other epochs, a standard Kalman filter has been considered due to the fact that allows to estimate not only the coordinates of the centre of view but also velocities and accelerations of the rover. In order to improve the quality of the solution, an outlier rejection technique has been considered based on a data snooping approach. This method is based on the "test of the normalized residual", applied to normal variables, with average $M = 0$ and variance $\sigma^2 = 1$.

In the case of uncorrelated observations, residuals w to be tested are the normalized residuals, divided by their standard deviation:

$$w_i = \frac{v_i}{\sigma_{v_i}} \quad \text{with} \quad w_i \in N(0,1) \quad (4)$$

If the normalized residual exceeds the normalized confidence interval $|w_i| > k_\alpha$, with a significance level α and a confidence interval k_α , it means that we are in a presence of an outlier. So, after the detection it is important to remove that from the system of equations: considering the Kalman filter, this can be easily removed deleting the equation without rewriting the whole system, allowing the real-time estimation of the positioning solution.

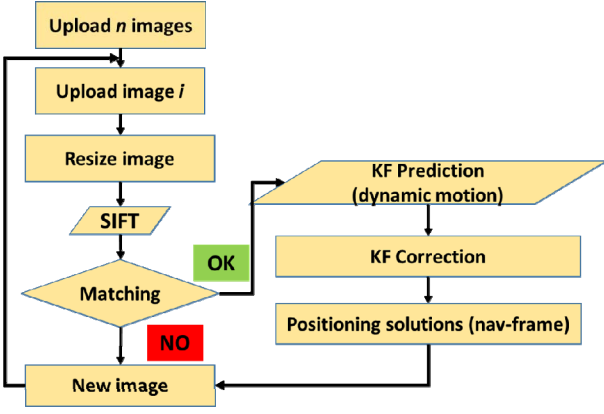


Fig. 6. The flowchart of the proposed procedure

As it is well-known, the acquisition phase is difficult as the definition of the ground truth, especially for underwater environments: in this case, the authors have considered the solution provided by the owners of the AQUALOC dataset as reference, that was obtained considering the state-of-the-art structure-from-motion (SfM) library Colmap [24] to compute comparative baseline trajectories for each sequence. This method computes a post-processing 3D reconstruction to estimate the position of the cameras by matching all the collected images, composing a sequence and exploiting the loop-closure information. Finally, it is possible to analyze the obtained trajectory as well as the reprojection errors and the statistics of the 3D reconstruction for evaluating their accuracy.

IV. RESULTS

As described in the previous section, the AQUALOC dataset has been used for testing the proposed algorithm. Considering the positioning results, the attention has been firstly focused on the prediction of the estimation: in order to strength the solution using a Kalman Filter, it has been decided to predict the new position of the camera at epoch t in function of the position evaluated at epoch $t-1$. This has been done considering Eq. (1), assuming that velocity and acceleration of the rover can be estimated by the Eq. (2) and (3). An alternative solution could be represented by the velocity and acceleration obtained directly from the UUV or from the IMU: if in the first case it is not possible due to the impossibility of extracting this information, the second case will be investigated in the future. Despite that, the obtained results are good, as possible to see from Table 5: the maximum difference between predicted and estimated positions are about 8 cm for the 3D components.

TABLE V. COMPARISON OF FEATURE DETECTORS: PROPERTIES AND PERFORMANCE

| Indicators | Results |
|------------------------------|---------|
| Number of used images | 630 |
| Number of 3D points | 248,036 |
| Mean reprojection error (px) | 0.673 |

| | |
|----------------------|-------|
| Maximum 3D error [m] | 0.078 |
| Mean 3D error [m] | 0.023 |

Even from a graphical point of view, the estimated trajectory seems comparable with the reference one (Fig. 7): the main differences can be found at the beginning of the survey, where the ROV is in a pseudo-static situation. This is more visible if the up component is analyzed, as shown in Fig. 8. This problem could be probably overcome in two possible ways: or estimating better the features during the photogrammetric approach or performing a better tuning approach in the Kalman filter, considering for example an extended (EKF) or unscented (UKF) Kalman filter. These approaches will be investigated in the future steps of this research activity, as well as integrating other external sensors for improving the quality of these results both in terms of precision and accuracy.

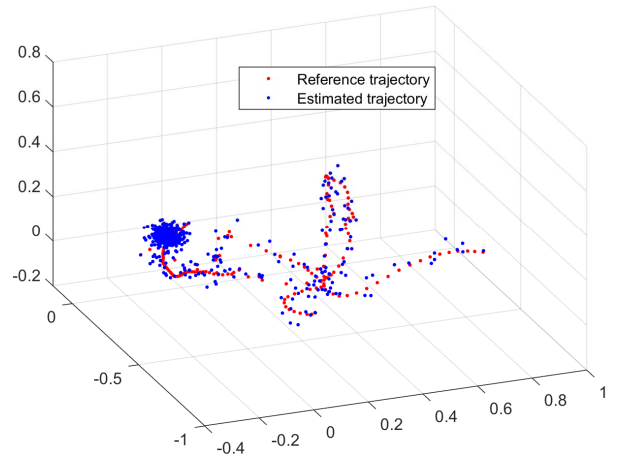


Fig. 7. Comparison between the estimated and reference 3D trajectories

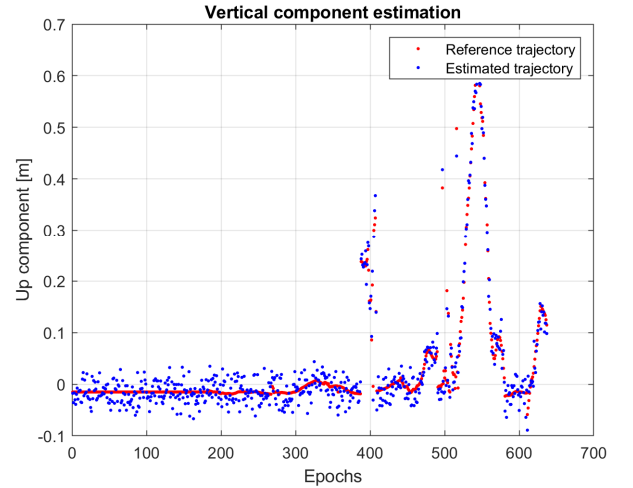


Fig. 8. Comparison between estimated and reference Up components

V. CONCLUSIONS

The importance of underwater positioning has increased in the last decade due to many purposes' requests, like

construction, communication, localization, control and deployment due to the nature of the environment. With these instruments, it is quite easy to perform visual inspections, maintenance and repair of dams, pipes, tunnels, structures as well as the analyses of the underwater environments in lakes, rivers and seas. Many different UUVs, in terms of architectures and sensors, have been developed for reaching these goals but the main issue is the determination of an accurate geo-referenced positioning solution every time and everywhere. In the present work a positioning solution based on monocular visual odometry has been proposed, using an open-source dataset available online, without the use of any other external device, such as IMUs or pressure sensors.

The obtained results have shown an impressive performance of the algorithm developed, reaching a maximum difference of about 8 cm with respect to the reference solution, obtained using a classical photogrammetric approach, based on SfM. The maximum error has been obtained at the beginning of the survey, where the ROV was in a static condition. Thus, a better dynamic motion in the prediction phase of the Kalman filter algorithm could be considered, in order to reduce the noise of the results. Another possible solution could be the integration of other external information, like the accelerations and angular velocities extracted from the IMU platform, in order to implement a loosely or tightly coupled integration.

This work is the first step of a new research line of the PIC4SeR Interdepartmental Research Centre at Politecnico di Torino, where a couple of UUV are now available for study and experimentation of systems, instruments and innovative methodologies for the survey and inspection of infrastructures in complex environments. In future work, we plan to perform new acquisition missions in different underwater environments in order to improve the results, consider other environments, create and share new datasets for increasing the development and improvement of algorithms.

ACKNOWLEDGMENT

All authors want to thank the owners of the AQUALOC dataset for sharing this interesting dataset.

The present work is supported by the PIC4SeR center at Politecnico di Torino, which all authors are belonging.

REFERENCES

- [1] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1980.
- [2] Paull L, Saeedi S, Seto M and Li H (2014) AUV navigation and localization: A review. *IEEE Journal of Oceanic Engineering* 39(1): 131–149.
- [3] Cadena C, Carlone L, Carrillo H, et al. (2016) Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* 32(6): 1309–1332.
- [4] Dabove P, Lingua AM and Piras M (2018) Photogrammetric visual odometry with unmanned ground vehicle using low cost sensors. In 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS) (pp. 426–431). IEEE.
- [5] Engel J, Koltun V and Cremers D (2018) Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(3): 611–625
- [6] Forster C, Zhang Z, Gassner M, Werlberger M and Scaramuzza D (2017) SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics* 33(2): 249–265
- [7] Mur-Artal R, Montiel JMM and Tardos JD (2015) ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics* 31(5): 1147–1163.
- [8] Leutenegger S, Lynen S, Bosse M, Siegwart R and Furgale P (2015) Keyframe-based visual-inertial odometry using non-linear optimization. *The International Journal of Robotics Research* 34(3): 314–334
- [9] Mur-Artal R and Tardos JD (2017) Visual-Inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters* 2(2): 796–803
- [10] Qin T, Li P and Shen S (2018) VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* 34(4): 1004–1020
- [11] Geiger A, Lenz P and Urtasun R (2012) Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, pp. 3354–3361
- [12] Blanco JL, Moreno FA and Gonzalez-Jimenez J (2014) The Malaga Urban Dataset: High-rate stereo and lidars in a realistic urban scenario. *The International Journal of Robotics Research* 33(2): 207–214
- [13] Burri M, Nikolic J, Gohl P, et al. (2016) The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163
- [14] Duarte, AC, Zaffari, GB, da Rosa, RTS, Longaray, LM, Drews, P, Botelho, SSC (2016) Towards comparison of underwater SLAM methods: An open dataset collection. In: OCEANS 2016 MTS/IEEE, Monterey, CA, USA, pp. 1–5.
- [15] Mallios, A, Vidal, E, Campos, R, Carreras, M (2017) Underwater caves sonar data set. *The International Journal of Robotics Research* 36(12): 1247–1251
- [16] Ferrera M, Creuze V, Moras J and Trouvé-Peloux P (2019) AQUALOC: An underwater dataset for visual-inertial-pressure localization. *The International Journal of Robotics Research*, 38(14), 1549–1559.
- [17] Burguera A, Bonin-Font F and Oliver G (2015) Trajectory-based visual localization in underwater surveying missions. *Sensors* 15(1): 1708–1735.
- [18] Ferrera M, Moras J, Trouvé-Peloux P and Creuze V (2019) Real-time monocular visual odometry for turbid and dynamic underwater environments. *Sensors* 19(3): E687
- [19] Creuze V (2017) Monocular odometry for underwater vehicles with online estimation of the scale factor. In: IFAC 2017 World Congress, Toulouse, France.
- [20] Rahman S, Li AQ and Rekleitis I (2018) Sonar visual inertial SLAM of underwater structures. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, pp. 1–7
- [21] Piras M, Lingua AM, Dabove P, and Aicardi I (2014) Indoor navigation using Smartphone technology: A future challenge or an actual possibility?. In 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014 (pp. 1343–1352). IEEE.
- [22] Lingua A, Marenchino D, and Nex F (2009) Performance analysis of the SIFT operator for automatic feature extraction and matching in photogrammetric applications. *Sensors*, 9(5), 3745–3766.
- [23] Masiero A (2017) Photogrammetric 3d reconstruction in Matlab: development of a free tool. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.
- [24] Schonberger JL and Frahm JM (2016) Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA.
- [25] Gonzalez R, and Dabove P (2019). Performance Assessment of an Ultra Low-Cost Inertial Measurement Unit for Ground Vehicle Navigation. *Sensors*, 19(18), 3865.

- [26] IEEE-SA Standards Board. IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros; IEEE Std 952TM-1997 (R2008); IEEE: New York, NY, USA, 1998.
- [27] Allan DW (1966) Statistics of atomic frequency standards. *Proc. IEEE* 1966, 54, 221–230.
- [28] El-Sheimy N, Hou H, and Niu X (2008) Analysis and modeling of inertial sensors using Allan variance. *IEEE Trans. Instrum. Meas.* 2008, 57, 140–149
- [29] Forstner W (1986) A feature based correspondence algorithm for image matching. *Int. Arch. Photogrammetry*, vol. 26, no. 3, pp. 150–166.
- [30] Harris C, and Pike J (1988) 3D positional integration from image sequences, in *Proc. Alvey Vision Conf.*, pp. 87–90.
- [31] Tomasi C, and Shi J (1994) Good features to track,” in *Proc. Computer Vision and Pattern Recognition (CVPR '94)*, pp. 593–600.
- [32] Rosten E, and Drummond T (2006) Machine learning for high speed corner detection, in *9th European Conference on Computer Vision*, vol. 1, 2006, pp. 430–443.
- [33] Bay H, Ess A, Tuytelaars T, and Van Gool L (2008) SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp. 346–359.
- [34] Milella A, and Siegwart R (2006) Stereo-based ego-motion estimation using pixel tracking and iterative closest point, in *Proc. IEEE Int. Conf. Vision Systems*, pp. 21–24.
- [35] Milford M, Wyeth G, and Prasser D (2004) RatSLAM: A hippocampal model for simultaneous localization and mapping, in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA '04)*, pp. 403–408..